Defense Technical Information Center
Compilation Part Notice

This paper is a part of the following report:

• *Title:* Technology Showcase: Integrated Monitoring, Diagnostics and Failure Prevention. Proceedings of a Joint Conference, Mobile, Alabama, April 22-26, 1996.

• *To order the complete compilation report, use:* AD-A325 558

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

HELICOPTER TRANSMISSION DIAGNOSTICS USING VIBRATION SIGNATURE ANALYSIS

B.E. Parker, Jr., M.P. Carley, R.J. Ryan, and M.J. Szabo
BARRON ASSOCIATES, INC.
3046A Berkmar Drive
Charlottesville, Virginia 22901-1444

and

H.V. Poor
PRINCETON UNIVERSITY
Department of Electrical Engineering
Princeton, New Jersey 08544-5263

Abstract: Fault diagnostics represent a vital task in the monitoring of mission critical systems, as well as for condition-based maintenance of machinery in general. The focus of this report is on the early detection, and subsequent classification, of small changes in the behavior of mechanical systems. Such changes, known as *incipient* faults, portend the development of more serious failures.

Physical models of machinery processes, which are useful for model-based fault detection and isolation, are not generally available in most applications. Instead, the approach to fault detection considered in this study involves the application of *statistical change detection.* Statistical change detection is essentially the problem of homogeneity testing within a time series. In particular, statistical change detection algorithms seek to detect situations in which a given model that describes the initial behavior of a time series, eventually fails to describe that time series accurately. The performance of non-likelihood-ratio techniques are evaluated on a CH-47D helicopter combiner transmission (non-seeded) fault; results indicate that the fault is detected in its incipient stage.

The approach to fault *isolation* (i.e., classification) discussed herein is based on the use of minimum-logistic-loss polynomial neural networks (PNNs). The fault isolation capabilities of PNN classification networks are investigated using seeded-fault data taken from CH-46E helicopter combiner transmissions. Perfect fault classification results are achieved.

Key Words: Condition-based maintenance, Fault diagnostics, Health monitoring, Incipient faults, Nonparametric statistics, Pattern recognition, Polynomial neural networks, Statistical change detection

Introduction: Mechanical system maintenance is inherently a safety issue for critical systems such as helicopter transmissions, since even a single in-service failure is intolerable. The maintenance strategy in such systems is made more difficult by the need to minimize unnecessary inspections, since this requires expensive teardown and rebuild procedures. Alternatively, it is also expensive to have unexpected unavailability of such systems. What is needed to address this dilemma in commercial, industrial, and military systems, such as aviation and manufacturing process control, is transition from a "time-based" to "condition-based" maintenance (CBM) strategy. Such a transition is critical because of the costs associated with maintenance. Enterprises employing such diagnostic capabilities would realize enhanced responsiveness, competitiveness, profitability, and consumer image. The need for diagnostics capabilities is, if anything, even greater for military systems, since maintenance requirements during military conflict impart a tactical disadvantage. At such times, spare aircraft, for example, may simply not be available.

In the case of helicopters, on-board health and usage monitoring systems (HUMS) could be integrated with groundstations, analyzing data that are collected and stored for post-flight analysis. Affordable computer workstations could readily provide the needed processing power (special-purpose *hardware* is *not* needed). For most mechanical system faults, *need* for an on-line (i.e., airborne) system would be determined by the possible risk of the off-line (i.e., groundstation) system not detecting a fault sufficiently early for the operator to take maintenance action, or not having significant ground support available (such as the situation of forward deployment). Rare faults can initiate and progress to complete failure within the time of a single mission (e.g., < 4 hours), so a trade-off must be made between the costs and benefits of on-board (on-line) and post-flight (off-line) analyses.

The goal of this work is to develop capabilities that will lead to affordable systems that can diagnose faults sufficiently early to significantly enhance safety, improve the likelihood of accomplishing mission objectives, prevent the loss of assets, and reduce maintenance costs.

Diagnostics: Model-based fault detection and isolation (FDI) is the method of choice when physical models of machinery processes are available (see, e.g., [6]). With the present state of the art, however, such models are available only for very simple and idealized mechanisms, which do not capture the necessary complexities of real-world processes. Even relatively simple gearboxes, such as those of helicopter tail rotors, can have a dozen or more shafts when auxiliary equipment is included, along with dozens of gears and bearings.

Although pattern-recognition techniques are often used to implement fault detectors via classifiers that process feature-set data, such an approach is not entirely satisfactory. First, it is not possible to ensure with *ad hoc* feature selection that one has chosen a sufficient feature set with which to distinguish normal behavior from all possible abnormal behaviors. Even if the features selected perform well on the available training and test data (a necessary condition), this provides no information regarding the sufficiency of such features for detecting previously unseen types of faults. Validation of the feature set using available data does not ensure that abnormal data will not be encountered that the system is unable to classify correctly.

Second, pattern-recognition techniques used for detection that rely on extracted features often require access to a significant amount of training data to achieve acceptable performance. In many applications such data are expensive to acquire (e.g., flight vehicles) and may therefore be available only in very limited quantities. Additionally, feature-based approaches require an inordinate amount of labor for designing each detector/classifier system, and it is unlikely that features engineered for one system can be used reliably in other systems of different design. Even though the same features may, in part, be useful, the entire process of validating the features on test data has to be repeated for each such system, once again with no real assurances regarding their sufficiency.

A better approach, both from the viewpoint of increasing the reliability of detection and in reducing development costs (through reduced engineering manpower needed to synthesize such detectors), is to use a general methodology that can be applied more readily to different systems. *Statistical change detection* represents such an approach that can be used for the early detection of small changes in systems. Statistical change detection does *not* require a database of fault examples, allowing *novel* situations and faults to be detected. Following fault detection, fault classification may be performed by utilizing pattern recognition techniques (e.g., neural networks). With fault classification, "feature engineering" also may be avoided by using parameters of the models employed in the statistical change detection algorithms as "features;" these parameters would need to be adapted on line so as to continually "fit" the data. (Note that the detector would have to be isolated from such adaptation, since the detector relies on a fixed whitening model — even if all of its coefficients are set to zero.) The motivation for such use of the parameters is that it obviates the need to seek *ad hoc* feature sets for fault classification in machinery, which is the approach currently taken in nearly all diagnostic systems that are not based on parametric models. The justification for such use of these parameter values is that many modern spectral analysis techniques are based on nonparametric (e.g., autoregressive, moving average, autoregressive-moving average, etc.) models; these techniques utilize model parameter estimates to obtain spectral densities.

Statistical Change Detection: Generally speaking, the detection algorithms consist of two parts: signal processing to reduce the observed times series data to a single real number, i.e., the computation of a *detection statistic*; and the comparison of this real number to a threshold to make a decision. The signal processing can, in principle, be implemented with no *a priori* data, by synthesizing whitening filters (both their structure and parameter values) on-line. The use of neural network algorithms can be particularly beneficial here. The second part of detection requires determination of the detection threshold, for which access to some *a priori* data is generally needed, but in certain cases this threshold can be set using theoretical values that do not depend on data (e.g., use of χ^2 statistics); additionally, useful thresholds can also be learned adaptively on-line by monitoring false-alarm statistics. In monitoring mission critical systems for potential faults, it is vital that false alarms be minimized, since mission abort decisions have their own price (e.g., when aborting a mission over water or hostile territory).

Statistical change detection thus provides a framework within which vibration monitoring of mechanical systems can be approached. Discrete-time change detection problems can be formulated as parametric hypothesis testing problems based on a series $y_1, y_2, \ldots, y_n$ of random measurements. We assume that these measurements are generated by a statistical model M_0 up until an unknown time $t_0 - 1$, and then that they are generated by another statistical model M_1 thereafter.

There are two types of change-detection paradigms of interest: off- and on-line detection. In off-line change detection, the problem of interest is to examine a fixed-length set of measurements $y_1, y_2, \ldots, y_n$, and to decide whether or not $t_0 \leq n$. In this framework, the detection criterion is to maximize the probability of detecting a change, within a constraint on the false-alarm probability. In the on-line detection framework, the problem of interest is to continuously monitor the observations to detect the change point t_0 as quickly as possible after it occurs, again within a constraint on the allowable rate of false alarms.

Because of the similarity between off- and on-line change detection methodologies, the signal-processing algorithms used in the two approaches are similar structurally, and in fact can be identical if computational issues are not of concern. Where the algorithms differ is in the decision-making after the signal processing has been performed. This difference in decision-making procedures stems from the difference in their performance criteria: maximum detection sensitivity for off-line algorithms (maximization of the power of the test subject to the constraint of a fixed probability of false alarm), and quickest detection for on-line algorithms (minimization of the delay for detection for a given mean time between false alarms).

For general change detection, the detection statistic is based on a fixed-length sliding window of observations $y_{n-n_0}, y_{n-n_0+1}, \ldots, y_n$, and is a function of the form:

$$\Delta^n = \max_{n-n_1 \leq j \leq n-n_0} \Delta_j^n, \tag{1}$$

where, $n_1 - n_0 + 1$ is the length of the sliding window and n_0 is selected arbitrarily. The detection statistic Δ^n is suitable, in the general case, for deciding whether or not $t_0 \leq n$. For each $j = 1, 2, \ldots, n$, Δ_j^n is a suitable detection statistic for deciding between the hypothesis that the model changes from M_0 to M_1 at exactly time $t_0 = j$ versus the hypothesis that the model does not switch at all during the n observations (i.e., that $t_0 > n$).

The explicit structure of the detection statistic depends on the two models M_0 and M_1, on the nature of the difference between them, and on the complexity that can be tolerated by the detection system. In this context, there are four general types of detection statistics: the *log-likelihood ratio*, the *generalized likelihood ratio*, the *locally optimum* statistic, and *non-likelihood ratio (NLR)*-based statistics. These statistics were reviewed recently in [11, 12]; extensive comparisons of their performance were also provided via simulations. Emphasis herein will be placed on NLR-based detection statistics since these require the fewest assumptions; in particular, investigation is made into the Zhang, Basseville, and Benveniste (ZBB) algorithm [19, 12], and of the Basseville-Nikiforov (BN) algorithm [2, pp. 415 - 417].

After computation of the detection statistic from Eq. (1) through any of the above methods, the decision algorithms for change detection are quite simple. For off-line change detection, the presence of a change during the observation time window is announced only if $\Delta^n > \tau_{\text{off}}$, where τ_{off}

is a threshold set as small as possible while satisfying a false-alarm constraint, $P(\Delta^n > \tau_{\text{off}}|M_0) \leq \alpha$, where $P(\cdot|M_0)$ denotes probability computed under model M_0, and α is the desired constraint on the false-alarm probability. Alternatively, for on-line detection, a change from model M_0 to model M_1 is announced at the *alarm time*, t_a, given by $t_a = \min\{n|\Delta^n \geq \tau_{\text{on}}\}$, where τ_{on} is a decision threshold chosen to control the rate of false alarms. In general, the threshold in this process must be chosen to balance the desire for detection efficiency — i.e., a low threshold to achieve quick detection in on-line cases, high probability of detection in off-line cases — (which would indicate a low threshold) with the high threshold needed to minimize false alarms. The choice of threshold requires knowledge of the probability distribution of the statistic Δ^n under the model M_0. It is usually not possible to determine this distribution in most problems of vibration monitoring in mechanical systems; however, estimates based on Brownian motion approximations to the statistics Δ_j^n are available in most cases (see, e.g., [2, 19]). Useful thresholds can also be learned adaptively on-line by monitoring false-alarm statistics.

Nonparametric Models: In selecting a modeling approach, we are concerned with models whose function is to whiten (i.e., remove correlation from) the observed process (e.g., sensed vibration data), to produce an innovation or residual sequence that is used by the detection algorithm. For this purpose, *any modeling approach that is practical can be used to fit the data*; estimation neural networks are particularly useful here. The model should not, of course, overfit the data; moreover, parsimonious models are generally desired to allow fast on-line parameter estimation.

The consideration of external inputs as *nuisances* is paramount in the modeling of time-series data of systems where these inputs are varying, but for which no observability is provided (e.g., changing helicopter flight regime). *Instrumental variables (IV)* techniques exist that provide for parameter estimation under such circumstances (see, e.g., [7]). In particular, Basseville and Nikiforov [2] have shown that "the vibration monitoring problem is nothing but the problem of detecting and diagnosing changes in the eigenstructure of a nonstationary multivariable system in state-space form, or equivalently, in the AR part of a multivariable ARMA model with a nonstationary MA part." Because the AR parameters of systems are associated with its dynamics and the MA parameters with the (generally time-varying) input excitation, the latter are often treated as nuisance parameters and the AR process estimated using IV techniques. Instrumental variables techniques are especially important when reduced-order process models are used, as under-parametrization of models can result in situations where changes in the estimated system dynamics reflect only variations in conditions that are not being monitored [19].

It is also important to note that high-accuracy models are generally *not* needed for detecting incipient faults and small changes in machinery conditions; as a result, even non-harmonic processes, which are difficult to model well with a reasonably small number of parameters, can be monitored for purposes of detecting changes. Model consistency, at whatever level of accuracy, is the important attribute.

Second-order statistics are, in general, adequate when data being fitted are Gaussian. When data are non-Gaussian, or represent the output of a nonlinear process, higher-order statistics (HOS) may be more useful in characterizing a process. Part of the elegance of the statistical change detection approach is its ready extensibility to include higher-order statistical moments where necessary to exploit additional information in the data. In order to extend these methods to HOS, it is necessary only to produce a parametrized model based on HOS. One can then apply a generalization of the methodology used for second-order statistical estimation for the parameters of this model.

An essential element of the statistical change detection algorithms is that, although they too may exploit only second-order statistical information, they do *not* require reducing the "feature set" further. If higher-than-second-order statistics are to be used, feature-based methods grow even more problematic, for example, as bi-spectra now have on the order of N^2 spectral lines, and tri-spectra on the order of N^3 spectral lines, which must be reduced severely for inputting to a classifier.

Fault Isolation: To isolate (i.e., classify) faults, *classification* neural networks that employ a constrained minimum-logistic-loss criterion are most appropriate. The Barron Associates, Inc. *Algorithm for Synthesis of Polynomial Networks for Classification (CLASS)* [4] is used herein to synthesize these neural network classifiers. With *CLASS*, network outputs represent true estimates of the *a posteriori* probabilities of class membership.

Using an *estimation* neural network to perform classification is not optimal since it imposes unnecessary constraints on the solution [16]; for example, for binary classification, the network may be trained arbitrarily to output a "one" for a fault, and a "zero" for normal data. In essence, use of the squared-error loss function corresponds to the maximum likelihood rule only in the case of a Gaussian probability model for the distribution of the errors [7]. However, for multiclass classification problems with categorical variables, a multinomial probability model in regular exponential form is more suitable than the Gaussian model [1]. This approach is based on the estimation of nonparametric probability density functions using minimum-logistic-loss polynomial neural networks. The advantage of this approach is that the decision surfaces are more general and reflect distributions found in the data. In this case, the *CLASS* subnetwork (polynomial) functions are used to model the log-odds associated with the conditional probability of each class given the observed inputs. In this setting, the maximum likelihood rule corresponds to the choice of the logistic-loss function. Additionally, logistic discrimination has been shown to perform well on both Gaussian and non-Gaussian data. Another advantage of the general multiclass logistic model in regular exponential form is that, for classification problems, it forces satisfaction of the probability constraints $0 < p_i < 1$ and $\sum_i p_i = 1$.

In summary, the constrained minimum-logistic-loss criterion, which is explicitly designed for classification problems, provides performance superior to classifiers fitted using estimation criteria. Estimation networks place emphasis on estimation accuracy; minimum-logistic-loss networks instead place emphasis on maximizing the likelihood of correct class discrimination. Whereas true probabilities are always between 0 and 1, estimation networks are unbounded; in contrast, the logistic-loss criterion correctly maps the network outputs onto [0,1]. A significant advantage of the minimum-logistic-loss classifier is it gives the system a complete view of the problem at hand, with the coefficients in all nodes fitted simultaneously to the entire synthesis data set, instead of using separate fitting of partitioned data sets. This property also forces the trained nodes to be consistent with each other. The logistic-loss network is a completely general way of reflecting the natural distribution of the data, without imposing any assumed structure on the data.

Ensemble Processing for Cyclostationary Signals: The models discussed in the preceding sections were assumed to be stochastically stationary (within pre-change or post-change regimes); that is, their underlying statistical behavior was assumed to be invariant to arbitrary translations in time. However, some of the data sets considered in this study are not stochastically stationary, but rather they are *cyclostationary*, by which we mean that their underlying statistical behavior is invariant to time translations that are integral multiples of a basic time period, C. (See, e.g., [5] or [9] for a discussion of cyclostationarity.) For example, in the monitoring of a helicopter gearbox, cyclostationarity results from the cyclic motion of the gear shaft. In this section, we discuss modifications of the statistical change detection techniques for use on such cyclostationary signals.

It should be noted that a change in the statistics of a signal may not be manifested in the same way in every phase of the cyclic structure; evidence of a flaw in a gear, for example, may appear only in phases of the measurement signal during which the flawed part of the gear is engaged. Thus, processing the different phases of the signal together may reduce the detectability of such flaws. For this reason, it is of interest either to consider separate processing of the different phases of the cyclostationary signal, or to consider joint processing techniques that view the different phase signals as components of a C-dimensional vector time series.

In the first case, we can essentially treat each phase as an independent channel that can be processed by any of the scalar means described in the preceding sections. In this approach, the channels are combined only after per-channel processing has been performed. This combining can be either pre-decision or post-decision. For example, a detection statistic $\Delta^n(c)$ can be computed for each channel $c = 1, 2, \ldots, C$, and then threshold comparison can be performed with the combined statistic

$$\max_{1 \leq c \leq C} \Delta^n(c). \tag{2}$$

Alternatively, each per-channel statistic $\Delta^n(c)$ can be compared with a per-channel threshold λ_c, to produce C per-channel decisions, which can be combined to produce an overall decision. Such approaches will be designated *ensemble* approaches. The goal of ensemble processing is to improve fault detectability through signal-to-noise ratio enhancement resulting from coherent averaging. One such method in which a ZBB/NLR algorithm is used to perform per-channel processing is demonstrated below.

Boeing Helicopter Gearbox Data: The Boeing helicopter gearbox data analyzed in this report are based on vibration measurements performed on a CH-47D helicopter combiner transmission spiral bevel input pinion gear that exhibits a classic bending fatigue failure as a result of test-cell overloading (156 percent rated torque). Therefore, this data does *not* represent a seeded-fault test. Data collection techniques and other analyses performed previously on this data have been reported in [13].

The data available for this study were comprised of two accelerometer channels digitized from high-speed analog tape recordings at a sample rate of 121,212 Hz; one accelerometer was mounted on the combiner collector gear along an axis parallel to the output shaft center line. The other accelerometer was mounted on the lefthand combiner input pinion on an axis perpendicular to the input shaft center line. Both accelerometers had bandwidths of five to 20,000 Hz. A tachometer signal was also available that provided a pulse for each revolution of the gear experiencing the bending fatigue failure. The 23 Boeing data records available represent consecutive (but not contiguous) 30-second segments extracted from each minute of time-series data; these data segments initially represent normal operating conditions, and eventually include the initiation of a gear microcrack at the root of a tooth, and its progression to a complete tooth failure, at which time testing was terminated due to a sudden increase in the test cell external noise level.

Analysis Results Using Stewart Hughes Ltd. MSDA Analyzer: Based on metallurgical analyses of the gear tooth that fractured in this experiment, Boeing has determined that a microcrack initiated approximately 13.25 minutes prior to complete gear tooth failure and termination of the test data recording [13]. Alignment of the data available in the present study with that of the Boeing analyses indicates that the microcrack initiated midway in file #12.

Analysis of the same data by Boeing using the Stewart Hughes Ltd. Mechanical Systems Diagnostic Analyzer (MSDA) [13] led to post-fault initiation detection delays ranging from 2.85 minutes to 9.45 minutes (detection in files #15 and #22, respectively), depending on the failure indicator (i.e., figure-of-merit (FOM)) monitored. These time delays are based on the ten MSDA FOM, out of the 32 investigated, that showed positive indications (i.e., significant and continuous change from the long-term value developed during a period of normal operation) of the microcrack failure progressing with time.

Statistical Change Detection Using Time-Average Statistics: Because the data files are not contiguous segments of the original recording, an off-line change detection approach was used. Each of the available data files #1 - #23, with the exception of file #2, was subdivided into 139 segments of 25,000 observations each; data file #2 was not used in these analyses because one of the two digitized accelerometer channels was not readable by computer. Close examination of the raw data also revealed several sections of the second accelerometer (i.e., channel 7) time-series data that were saturated. In particular, files #6 and #20 contained regions in which the data appeared to be magnified in bursts relative to the majority of the recorded data. These sections were, therefore, eliminated. The first 25,000-sample segment of file #1, which represents normal operational data, was reserved for estimating parameters and training (i.e., learning the mean and covariance values used in computing the test statistics during evaluation).

Using the BN/NLR estimation algorithm [2], the AR parameters of a two-dimensional ARMA (2,1) process were found; the BN/NLR algorithm requires that the order of the MA process, q, be one less than the order of the AR process, p, thus $q = p - 1$. The MA parameter(s) are then not actually estimated, but suppressed using instrumental variables (IV) estimation techniques. Subsequently, all of the other segments of the two channels of data were evaluated individually based on the information provided by the training data sequence.

Based on the ratios of the mean values of the detection statistics, $\overline{\Delta^n}_{ev}/\overline{\Delta^n}_{tr}$ (where the subscripts ev and tr imply evaluation and training data respectively), which are provided graphically in Fig. 1, it is seen that the fault appears to be detectable first in file #20; it is detectable consistently thereafter (i.e., in files #21 - #23). Based on this result, the time duration between the estimated time of fault initiation (i.e., midpoint in file #12) and the time of detection (i.e., detection delay) was determined from the test time interval data to be approximately eight minutes.

Off-line change detection was also performed on the univariate (i.e., scalar) time-series data; these results are illustrated in Fig. 2. The results for the univariate algorithm are slightly worse than those for the multivariate algorithm in that the fault appears to be detectable consistently by file #21. The BN/NLR algorithm scalar detection results therefore indicate a detection delay of approximately nine minutes.

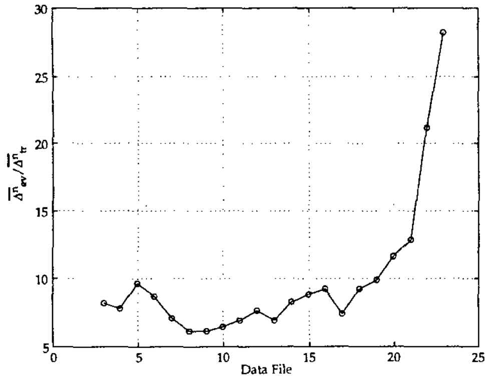

Figure 1: Off-Line Boeing Helicopter Gearbox Results (BN/NLR Multidimensional Algorithm)

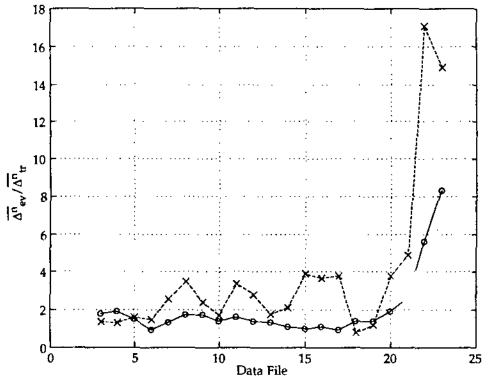

Figure 2: Off-Line Boeing Helicopter Gearbox Results (BN/NLR Algorithm); solid line represents accelerometer channel 3, dashed line accelerometer channel 7.

Statistical Change Detection Using Ensemble-Average Statistics: Analysis was performed also using the ZBB/NLR Ensemble algorithm. This algorithm differs from the time-average algorithm in that ensemble averages are used to calculate the detection statistics. The tachometer signal was used to divide the data into distinct records; l additional observation samples from the previous record were added to the beginning of each new record so that detection statistics could be calculated for the first samples of each record. These records were grouped into ensembles containing M records each. Breaking the data up in this way allows the basic statistics to be phase dependent, providing C different values, one for each phase of a record.

The use of ensemble averages is intended to help remove background noise, making the detection statistic more sensitive to change detection. Also, using a nonstationary bias and covariance matrix helps to improve modeling accuracy. A distinct detection threshold, λ_c, may be chosen for each phase within a record, further increasing the sensitivity of the detector.

Unfortunately, the time interval between tachometer pulses was not always constant, providing record lengths that varied slightly due to test stand speed fluctuations (+0.34% to −1.01%). To perform the off-line analysis, the two most common record lengths, 591 and 592 samples respectively, were selected and then decimated by a factor of two, to achieve equal record lengths of 296 samples (plus the l extra samples added to the beginning of each record). Use of this procedure allowed about 70 percent of all data records to be included in the analysis. More sophisticated techniques for synchronously averaging a signal are discussed in [14]. The decimated records from the first two files were grouped into ensembles containing $M = 50$ records. These ensembles were used, along with an AR(10) model, whose parameters were set to zero, to estimate the training parameters. The remaining data were then evaluated using ensembles containing $M = 100$ records.

The resulting statistics are graphed in Fig. 3. The ratio of the means of the test statistics for accelerometer channel 3 show a potentially significant change from the earlier statistics for data files #14 - #23. The fault was not as manifest on accelerometer channel 7, as the statistics do not show substantial change from a stable baseline until file #20. The results using the ensemble-average ZBB/NLR algorithm, therefore, provide for a best-case detection delay of two minutes, better than the other algorithms, including the MSDA analysis.

These results are based on use of the mean of the detection statistics computed over the entire record, which does not fully exploit the algorithm. A set of C thresholds, λ_c, one for each sample time (i.e., phase) within a record, were therefore established; each was arbitrarily set equal to the maximum value of the detection statistic seen in files #1 - #8. To avoid having to plot C different detection statistics, fault detection was based instead on the maximum number of consecutive times *any one* of the C phase-dependent detection statistics exceeded its corresponding λ_c. With this method, a significant change in the detection statistics may be seen as early as file #11, as shown in Fig. 4. This suggests that the microcrack may have initiated even earlier than the Boeing metallurgical analyses may have indicated, a finding that is not inconsistent with the approximate nature of such analyses [15].

For purposes of comparison, results were obtained next for the time-average ZBB/NLR algorithm using the same training and evaluation data and window lengths that were used to obtain the results for the ensemble-average ZBB/NLR algorithm. For training and evaluation the decimated data was used and M was set equal to 100. The ensemble-average ZBB/NLR algorithm (see Fig. 4) was found to detect the fault up to eleven minutes before the time-average ZBB/NLR algorithm. Thus, ensemble averaging clearly improves detection performance.

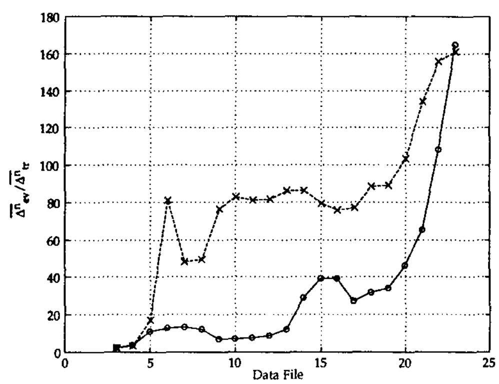

Figure 3: Off-Line Boeing Helicopter Gearbox Results (ZBB/NLR Ensemble Algorithm; solid line represents accelerometer channel 3, dashed line accelerometer channel 7.)

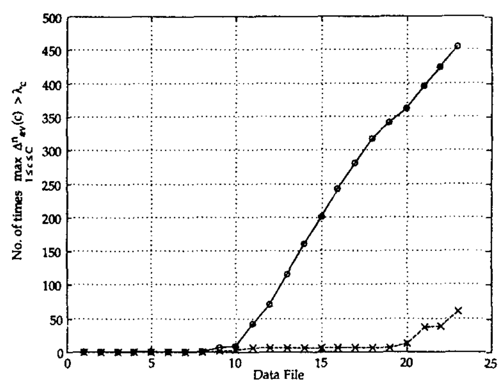

Figure 4: Off-Line Boeing Helicopter Gearbox Results (ZBB/NLR Ensemble Algorithm with Multiple Thresholds; solid line represents accelerometer channel 3, dashed line accelerometer channel 7.)

Fault Classification: Another important aspect of helicopter gearbox diagnostics is fault isolation (i.e., classification); in other words, determining the type of fault present once a fault has been detected. Contemporary fault classification, using neural networks, involves three basic steps. The first is feature extraction, in which pertinent features that can be used to distinguish one fault category from another are drawn from the data. A common method for extracting features is the short-term Fourier transform (STFT). Feature selection is the second step in classification. To simplify the neural network classifier and to avoid overfitting the data, the number of features input to the network must be kept to a minimum. Therefore, principal component analysis (PCA) [8] was used to reduce the dimension of the feature vector. The final step is synthesizing a classification neural network; here, the Barron Associates, Inc. *Algorithm for Synthesis of Polynomial Networks for Classification (CLASS)* [4] was used.

Classification Results for Westland Helicopter Gearbox Data: The Westland helicopter gearbox data analyzed in this report are part of a data set collected on the Westland Helicopter Universal Test Rig (see [17, 18]). The test rig was used to measure the vibration of a CH-46E helicopter combiner transmission under several different test conditions. These conditions were partitioned into nine different categories, as listed in Table I. Since "Fault Types" one and nine both represent "No Defect" cases, these categories were combined and are referenced as a single class (*viz.* Class 1).

Table I: Westland Helicopter Gearbox Data Description

Fault Type	Description
1	No Defect
2	Planetary Bearing Corrosion
3	Input Pinion Bearing Corrosion
4	S.B. Input Pinion Spalling
5	Helical Input Pinion Chipping
6	Helical Idler Gear Crack Prop.
7	Collector Gear Crack Prop.
8	Quill Shaft Crack Prop.
9	No Defect

Table II: Westland Helicopter Data Files, Fault Type vs. Torque Level

	Torque Level (%)									
Fault Class	27	40	45	50	60	70	75	80	100	Total
1	0	0	0	0	0	0	0	2	0	2
2	0	0	0	0	0	0	0	0	3	3
3	4	4	4	3	3	5	5	3	4	35
4	3	3	4	3	2	4	2	4	5	30
5	0	0	0	0	0	2	2	2	1	7
6	0	0	0	0	0	1	4	4	6	15
7	1	1	0	2	4	2	4	4	3	21
8	5	5	5	5	5	7	5	5	3	45
9	2	3	3	2	3	2	3	3	3	24
Total	15	16	16	15	17	23	25	27	28	182

Data were collected at different torque levels between 27 percent and 100 percent of the full torque level. In total, 182 different experiments were run. Table II shows the number of files obtained for each fault category as a function of the different torque levels. Each data file consists of approximately 22.6 seconds of data sampled at 100 kHz. Ten channels of data were recorded, which included eight accelerometers, one tach pulse, and one test signal. The tachometer was placed on the aft transmission in place of the rotor position motor. The tach signal is a 256 pulse-per-revolution signal w[illegible] a once-per-revolution signal superimposed on it. Based on its position in the gearbox, one revolution describes a complete rotation of the rotor position output, not that of the main shaft.

All of the available data files, except for file #24 on tape 1, were divided into one-revolution periods using the tachometer signal. File #24 was not utilized because it contained sections that were unreadable. With the 100 kHz sampling rate, there were between 897 and 904 samples within the period defined by the tachometer signal; all of these samples were used and then zero-padded to obtain 1,024 samples per period. Subsequently, 1,024-point power spectral density (PSD) estimates were computed for each period, after smoothing the data using a Kaiser-Bessel window. The PSDs were then clustered into groups of 20 and ensemble averaged. The resulting data vector of dimension 512, however, was too large for practical computation of the necessary PCA transformation matrix. Therefore, the PSDs were decimated by a factor of three to ease the computational burden; this resulted in reducing each PSD from 512 values to 179. The data were not decimated in the time domain to avoid the effects that this would have on the STFTs. Decimation by three in the time domain would cause loss of all information in the upper two-thirds of the STFT,[1] whereas decimation in the frequency domain maintained information at

[1]For example, with a 100 kHz sampling rate, spectral values are obtained from zero to 50 kHz; at a sampling

all frequencies, albeit at a courser resolution. The decimation operation is similar to averaging adjacent spectral lines to form a reduced set that still spans the original bandwidth. The resulting features were next normalized over the interval [0,1]. Lastly, the data were split into training and evaluation databases, each containing roughly half of the exemplars. To test the robustness of the classification algorithms, three different methods were used to effect this division; in the first approach, different data *files* were assigned randomly to the training and evaluation databases; in the second approach, data *exemplars* were apportioned randomly; in the third approach, data *exemplars* from the first half of individual time series were used for training, and exemplars from the second half were used for evaluation.

Since data were available at multiple torque levels within each fault category, a method was needed to reduce the within-class variance caused by changing torque levels, while simultaneously maintaining the between-class variance. Methods to accomplish this have been used before in sonar target recognition. In one study [3], a method was needed to reduce variation associated with the aspect angle of a target, while still being able to distinguish between different targets. In the present case, PCA was used to extract features within a class that tended to vary least with respect to torque level; these features are associated with the *smallest* eigenvalues. This process was used to find the 50 principal components within each class; these principal components account for the least amount of variance. Use of eight different one-vs.-all, binary-output neural networks, each specialized to recognize a single fault (or no-fault) class, were synthesized using *CLASS* [4] to provide a capability to distinguish between classes.

As discussed earlier, *CLASS* neural networks output estimates of the true *a posteriori* classification probabilities. To use the eight neural network outputs to reach a classification decision, the highest output probability was deemed to indicate the correct class, and the given exemplar was said to be of that class.

Use of any single accelerometer channel was found to lead to at least a few misclassifications. Therefore, all eight accelerometer channels were employed. To enable utilization of all accelerometer channels, eight binary-output neural networks were synthesized for each fault category (one for each accelerometer channel); the corresponding network output probabilities within each class were then averaged. This therefore required training of 64 different neural networks. A final fault category decision was reached based on the resulting probabilities in the same fashion as with the single channel method — i.e., that network whose output was largest was selected as winner.

Use of all accelerometer channels led to perfect classificatory results on both the training and evaluation data (where the latter were not used in training the classifier). Indeed, for *all* three partitionings of the data considered herein, interrogation of the neural networks with both training data and evaluation data exemplars always produced perfect classification results (i.e., diagonal confusion matrices). A typical confusion matrix, shown here for the random *file* partitioning, is provided in Table III.

Table III: Westland Helicopter Data Evaluation Confusion Matrix (Multiple Channels, Random File Partitioning)

	Model Output								
True Class	1	2	3	4	5	6	7	8	Total
1	1500	0	0	0	0	0	0	0	1500
2	0	250	0	0	0	0	0	0	250
3	0	0	2125	0	0	0	0	0	2125
4	0	0	0	1875	0	0	0	0	1875
5	0	0	0	0	375	0	0	0	375
6	0	0	0	0	0	875	0	0	875
7	0	0	0	0	0	0	1250	0	1250
8	0	0	0	0	0	0	0	2625	2625
Total	1500	250	2125	1875	375	875	1250	2625	10875

rate of 33.3 kHz, spectral values range from zero to 16.7 kHz.

It is noted that when there are only a few example data files for a given fault class, such as in the case of Class 2 (see Table II), classificatory results can be dependent somewhat on the assignment of particular files to the training or evaluation databases. For example, in the case of Fault Class 2 where there were only three fault examples, there are six possible ways in which the three files can be assigned to the training and evaluation databases (assuming that each database contains at least one file). In the case of Fault Class 2, it was found that two of the six possible partitionings led to perfect classification results on the evaluation data, whereas the other four partitionings resulted in a small number of misclassifications, which never exceeded 0.229% (14 out of 6,125 exemplar evaluations). Such results might be expected when so few data are available for training purposes.

Conclusions: The results documented in this report demonstrate success in two areas of fault diagnostics for helicopter transmissions: (1) rapid and accurate fault detection through the application of statistical change detection algorithms; and (2) perfect fault isolation (i.e., classification) using polynomial neural network classifiers synthesized with a constrained minimum logistic-loss criterion. Several examples of helicopter gearbox data, both seeded and non-seeded, were evaluated. The algorithms developed (i.e., statistical change detectors and trained neural network classifiers) are practical computationally, are data dependent rather than system dependent, and use nonparametric time-series models for change detection and polynomial neural networks for classification.

A vital characteristic of the non-likelihood ratio (NLR) fault detection algorithm is its robustness when utilizing reduced-order models. This is important in practice, as models used to monitor systems will necessarily be approximate. Although under-parametrization leads to some loss of information, if the model reduction is done well, detection of changes of interest can be achieved reliably, and all other changes can be treated as nuisances [19]. This provides a capability that is important in many practical applications (e.g., helicopter rotor transmission health monitoring). Inadequacy in the whitening model then becomes a secondary issue, affecting mainly detection *efficiency*. The use of neural networks is expected to play a significant role in the practical and flexible application of statistical change detection techniques. Here neural networks may be used as estimators to enhance detection when processing nonlinear and/or non-Gaussian data, and to automate the syntheses of both linear and nonlinear detectors.

Other desirable attributes of the NLR fault detection algorithm are its capabilities for quickest possible change detection (on-line algorithm) and maximum detection sensitivity (off-line algorithm) when the true models are *unknown*. The methodology is sufficiently general to be applicable across mechanical system designs without significant re-engineering effort. The algorithm does *not* rely on *ad hoc* feature characteristics, acting instead as a novelty detector, thereby obviating the need to collect examples of possible fault signature patterns for training purposes. The algorithm should enable more automatic syntheses of change detectors, placing the design burden on the adaptive algorithm, rather than the human. Without such automation, it is unlikely that the huge cost-savings potential of condition-based maintenance can be captured.

Best results were obtained with use of the ZBB/NLR multivariate algorithm and with ensemble-averag-ed, rather than time-averaged, detection statistics; the former are applicable whenever an adequate shaft-rate signal, such as a once-per-revolution tachometer pulse, is available (or can be derived [14]). Fault detection results on the CH-47D helicopter data are especially promising, because little attempt was made to use an optimal model structure to capture appropriate feature characteristics of the data or to whiten the data. In terms of the former, a simple AR(10) model was used, causing the algorithm to rely on the first ten autocorrelation coefficients for detection. In terms of the latter, the AR(10) model parameters were all set to zero, eliminating data whitening entirely.

Based on analyses of other data not reported on herein (see [10, 12]), it is likely that the difference statistically between different "normal" machines will be as large as those between normal and faulted machines, at least where faults in the latter case are incipient. This suggests that the detector may need to be "trained" on data taken from the particular machine in which changes are to be detected.

The excellent results obtained for fault classification also attest to the robustness of logistic-loss classifiers, since no special effort was required to achieve the demonstrated results. (Indeed, only one approach to data pre-processing and to training of a classifier was attempted for the CH-46E helicopter gearbox data; still, perfect *single-look* fault classification results were obtained!) This suggests that the polynomial neural network classifiers are also likely to perform well with other input features, including parameters obtained on-line through adaptation of the fault detection model.

References:

[1] A.R. Barron, "Statistical properties of artificial neural networks," *Proc. IEEE Conf. on Decision and Control*, Tampa, FL, December 1989.

[2] M. Basseville and I.V. Nikiforov, *Detection of Abrupt Changes*, 1993. (Prentice-Hall: Englewood Cliffs, NJ)

[3] P.C. Chestnut and R.W. Floyd, "An aspect-independent sonar target recognition method," *J. Acoust. Soc. Am.*, Vol. 70, No. 3, September 1981.

[4] *CLASS: Algorithm for Synthesis of Polynomial Networks for Classification Users' Manual*, Version 1.61, June 1995, Barron Associates, Inc.

[5] W.A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Magazine*, Vol. 8, No. 2, pp. 14 - 37, April 1991.

[6] C.J. Li and T. Kim, "Linear model-based fault detection and isolation for a screw compressor," *Mechanical Systems and Signal Processing*, Vol. 8, No. 3, 1993, pp. 259 - 273.

[7] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*, 1983, p. 4. (MIT Press: Cambridge, MA)

[8] K.V. Mardia, J.T. Kent, and J.M. Bibby, *Multivariate Analysis*, 1980. (Academic Press: London)

[9] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 1984. (McGraw-Hill: New York)

[10] B.E. Parker, Jr., T.M. Nigro, and M.P. Carley, *Machinery Diagnostics Using Vibration Signature Processing*, Barron Associates, Inc. Final Technical Rept. for Office of Naval Research, Contract N00014-92-C-0060, November 1992.

[11] B.E. Parker, Jr., H.V. Poor, M.P. Carley, N.A. Nigro, and R.J. Ryan, "Detection of small changes in mechanical systems," *Proc. of the U.S. Navy Small Business Innovation Research Conf.*, Arlington, VA, June 16 - 17, 1994, pp. 93 - 107.

[12] B.E. Parker, Jr., H.V. Poor, M.P. Carley, R.J. Ryan, M.J. Szabo, J.W. Stayman, N.A. Nigro, and P.D. Mooney, *Helicopter Transmission Diagnostics Using Vibration Signature Analysis*, Barron Associates, Inc. Final Technical Rept. for Office of Naval Research, Contract N00014-93-C-0074, July 1995.

[13] H.J. Rose, "Vibration signature and fatigue crack growth analysis of a gear tooth bending fatigue failure," *Current Practices and Trends in Mechanical Failure Prevention — Proc. 44th Meeting of the MFPG*, Virginia Beach, VA, April 3 - 5, 1990. (Vibration Institute: Willowbrook, IL)

[14] G.P. Succi, "Synchronous averaging of multiple gear shafts using an interpolating filter," *Life Extension of Aging Machinery and Structures — Proc. 49th Meeting of the MFPT*, Virginia Beach, April 18 - 20, 1995, pp. 113 - 125. (Vibration Institute: Willowbrook, IL)

[15] R.S. Teal, Boeing Defense & Space Group, personal communication.

[16] D.G. Ward, "Generalized networks for complex function modeling," *Proc. IEEE Systems, Man, and Cybernetics Conf.*, October 2 - 5, 1994, San Antonio, TX.

[17] Westland Helicopter, Ltd. Research Paper RP907, *Final Report on CH-46 Aft Transmission Seeded Fault Testing*, Vols. 1 and 2, September 1, 1993.

[18] Westland Helicopter, Ltd. Mechanical Research Report MRR20168, *CH-46 Aft Transmission Seeded-Fault Testing Analysis of Vibration Recordings*, January 1, 1994 (DTIC Document Number AD-B180 276).

[19] Q. Zhang, M. Basseville, and A. Benveniste, "Early warning of slight changes in systems," *Automatica*, Vol. 30, No. 1, 1994, pp. 95 - 113.

Acknowledgments: This investigation was funded by the Office of Naval Research under Contract N00014-93-C-0074. The authors thank Dr. Thomas M. McKenna for his kind support of this work. Gratitude is also expressed to Natalie A. Nigro of Barron Associates, Inc. for producing the figures contained herein, and to Susan L. Woodson, also of Barron Associates, for her technical typing assistance.